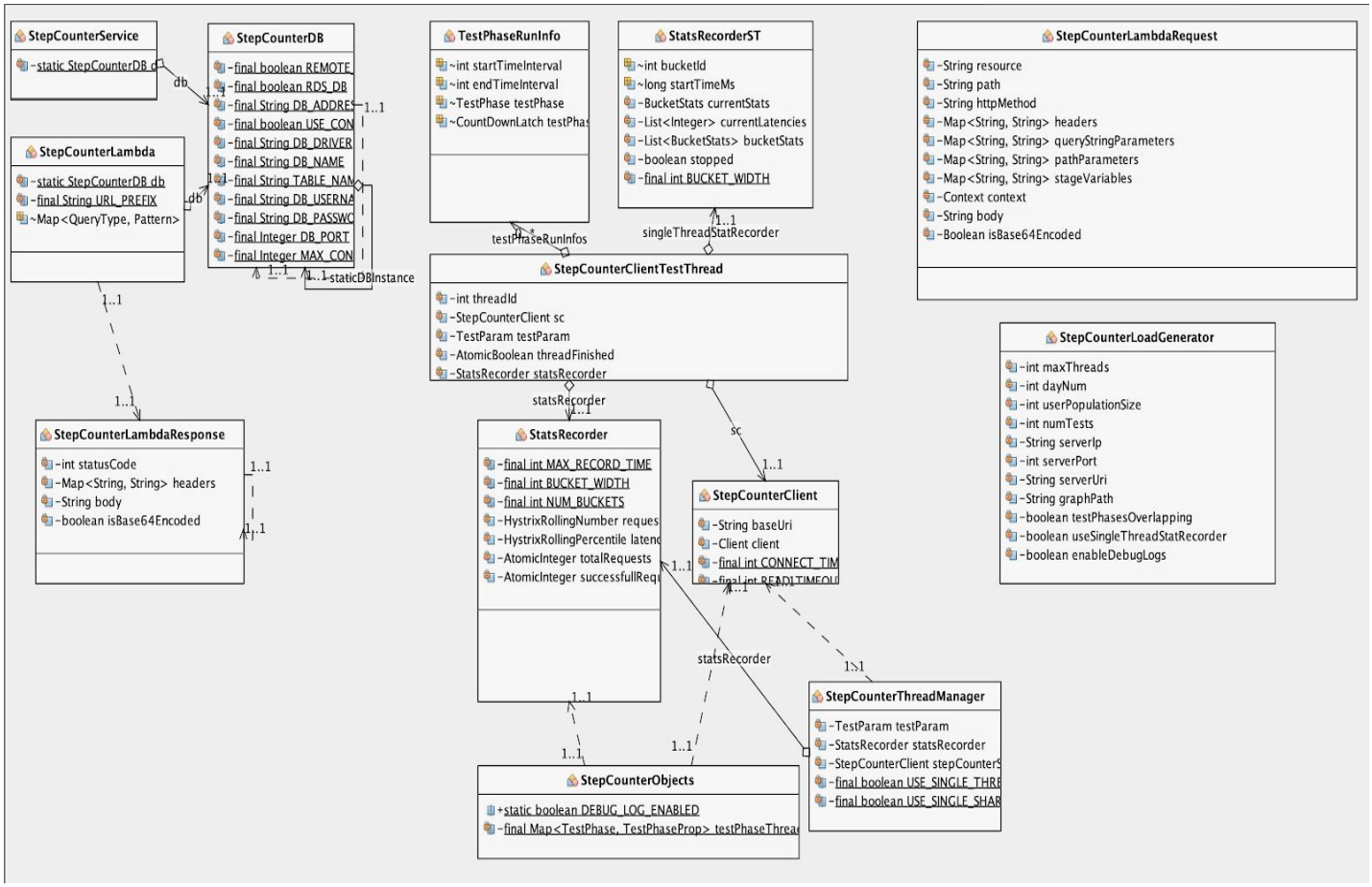
BSDS ASSIGNMENT 3

Submitted by: Manika Sharma

GITHUB REPO:

https://github.com/manika0407/bsdscourse/tree/master/Assignment3/bsdsassignment



STEPS 1

Step 1.1 **Lambda** 16, 32, 64, 128 Threads:

Lambda CMD :
I used Bash Script to run Lambda for 16, 32, 64, 128 threads.

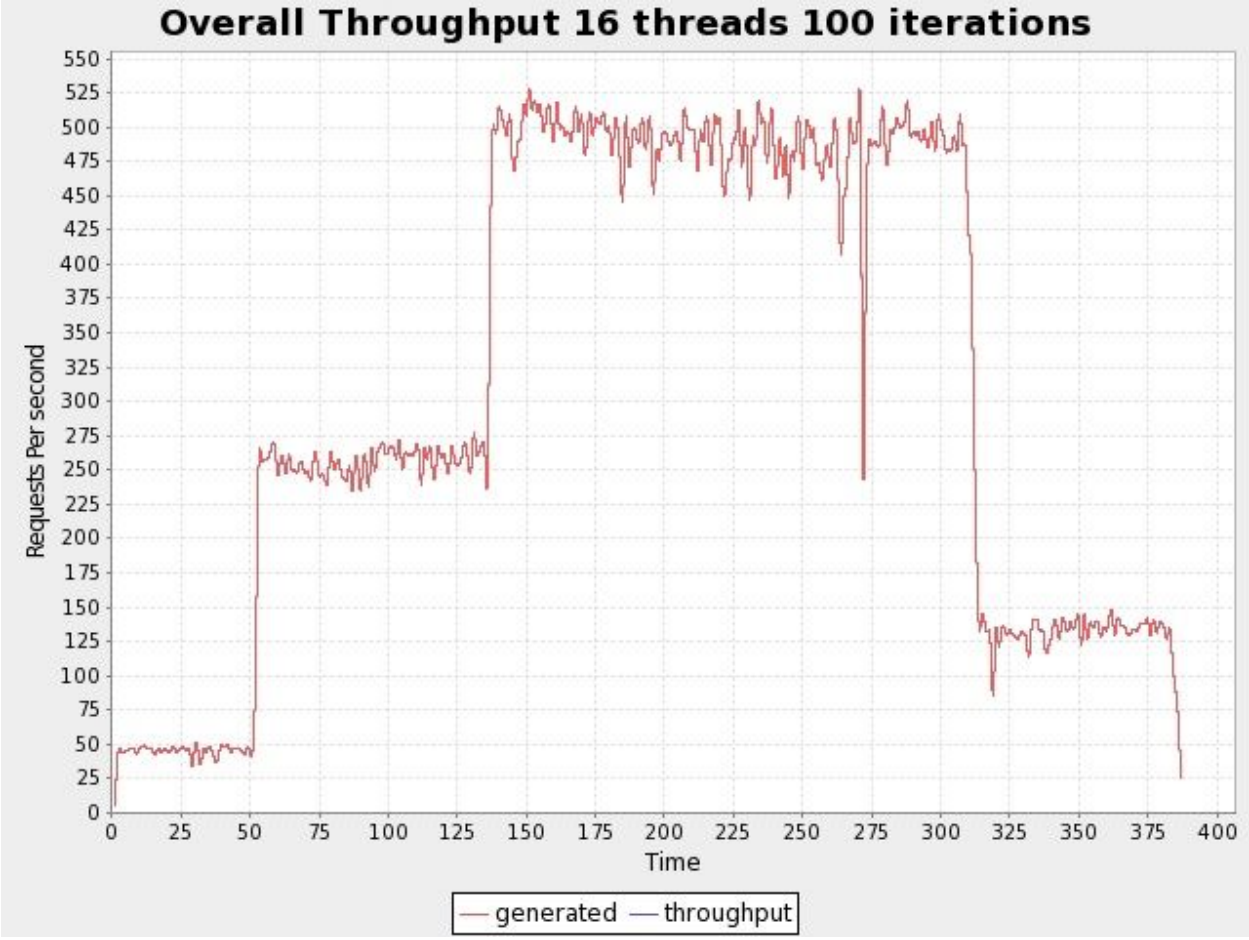
Script I used :

```
#!/bin/bash

for num_threads in 16 32 64 128 ; do
        echo "running for $num_threads threads"
```

```
        java -jar bsdsassignment2-stress-jar-with-dependencies.jar -h
"3fknpjdtrk.execute-api.us-west-2.amazonaws.com" -p 0 -u
"/prod/bsdsassignment2-webapp/stepcounter" -t "$num_threads" -g
"photos/StepCounterThroughput_lambda_${num_threads}.jpeg" 2>&1 | tee
"results/stats_lambda_${num_threads}.txt"
        echo "finished for $num_threads threads"
        sleep 60
Done
```

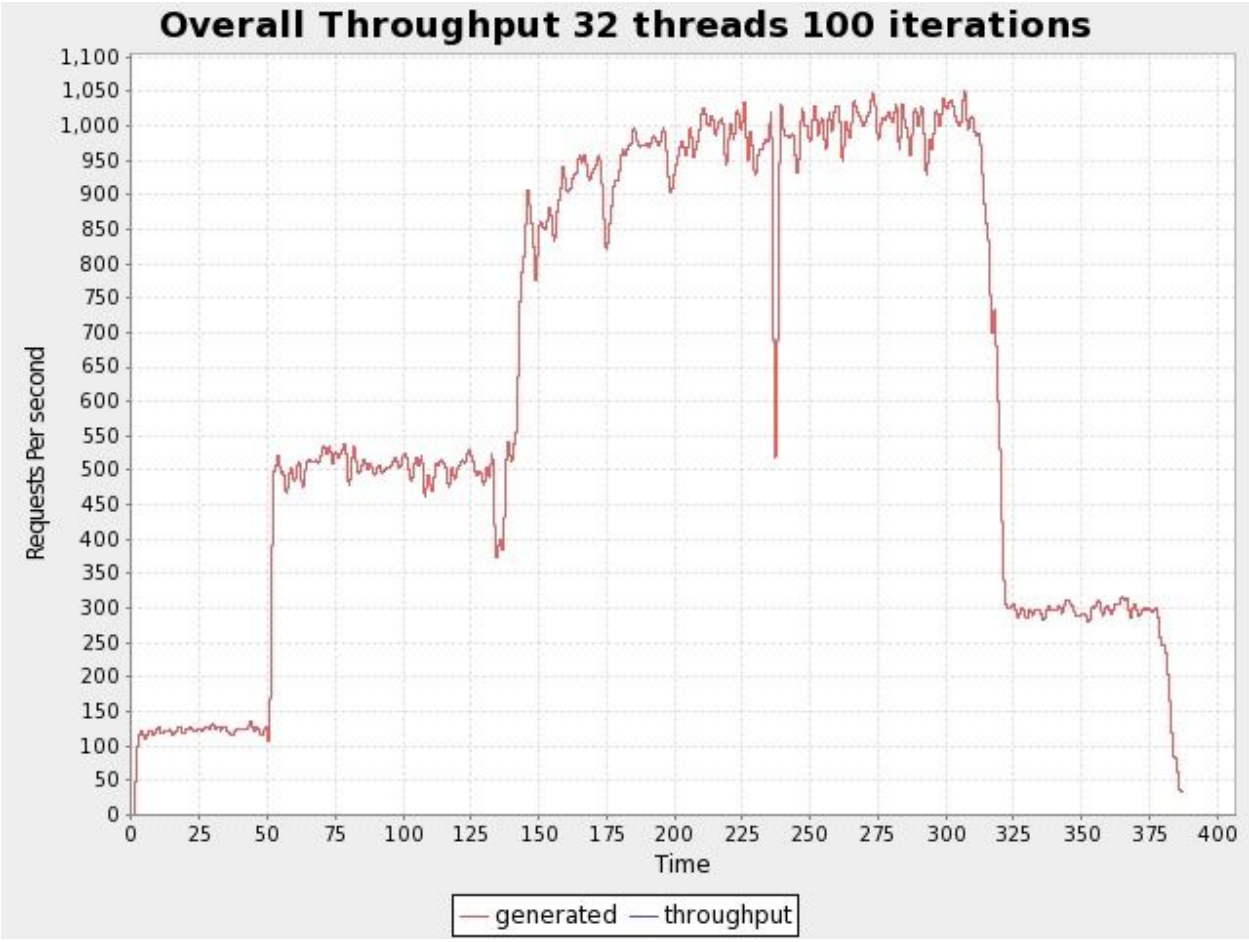
16 Threads :



Client starting.... Time: 1542244305015
WARMUP: All threads(1) running....
WARMUP complete: Time 49.524 seconds
LOADING: All threads(8) running....
LOADING complete: Time 82.36 seconds
PEAK: All threads(16) running....
PEAK complete: Time 180.533 seconds
COOLDOWN: All threads(4) running....

COOLDOWN complete: Time 73.815 seconds
==========
Total number of requests sent: 119500
Total number of Successful responses: 119500
Test Wall Time: 386.235 seconds
Overall throughput across all phases: 309.39712869108183 rps.
P95 Latency = 37 ms.
P99 Latency = 47 ms.

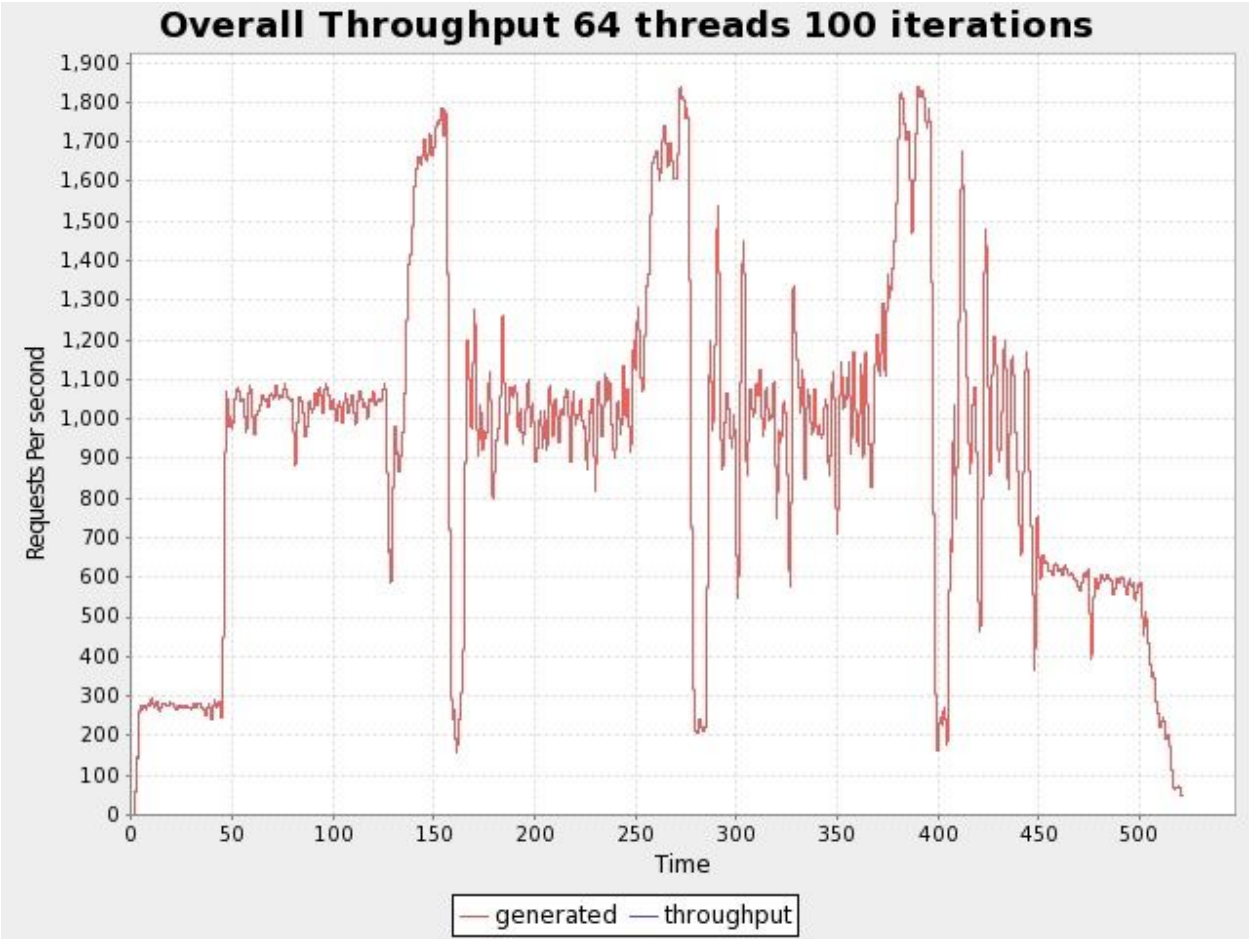
32 Threads:



Client starting.... Time: 1542244753892
WARMUP: All threads(3) running....
WARMUP complete: Time 48.476 seconds
LOADING: All threads(16) running....
LOADING complete: Time 83.7 seconds
PEAK: All threads(32) running....

PEAK complete: Time 189.234 seconds
COOLDOWN: All threads(8) running....
COOLDOWN complete: Time 64.309 seconds
==========
Total number of requests sent: 240500
Total number of Successful responses: 240500
Test Wall Time: 385.719 seconds
Overall throughput across all phases: 623.5108978297673 rps.
P95 Latency = 31 ms.
P99 Latency = 41 ms.

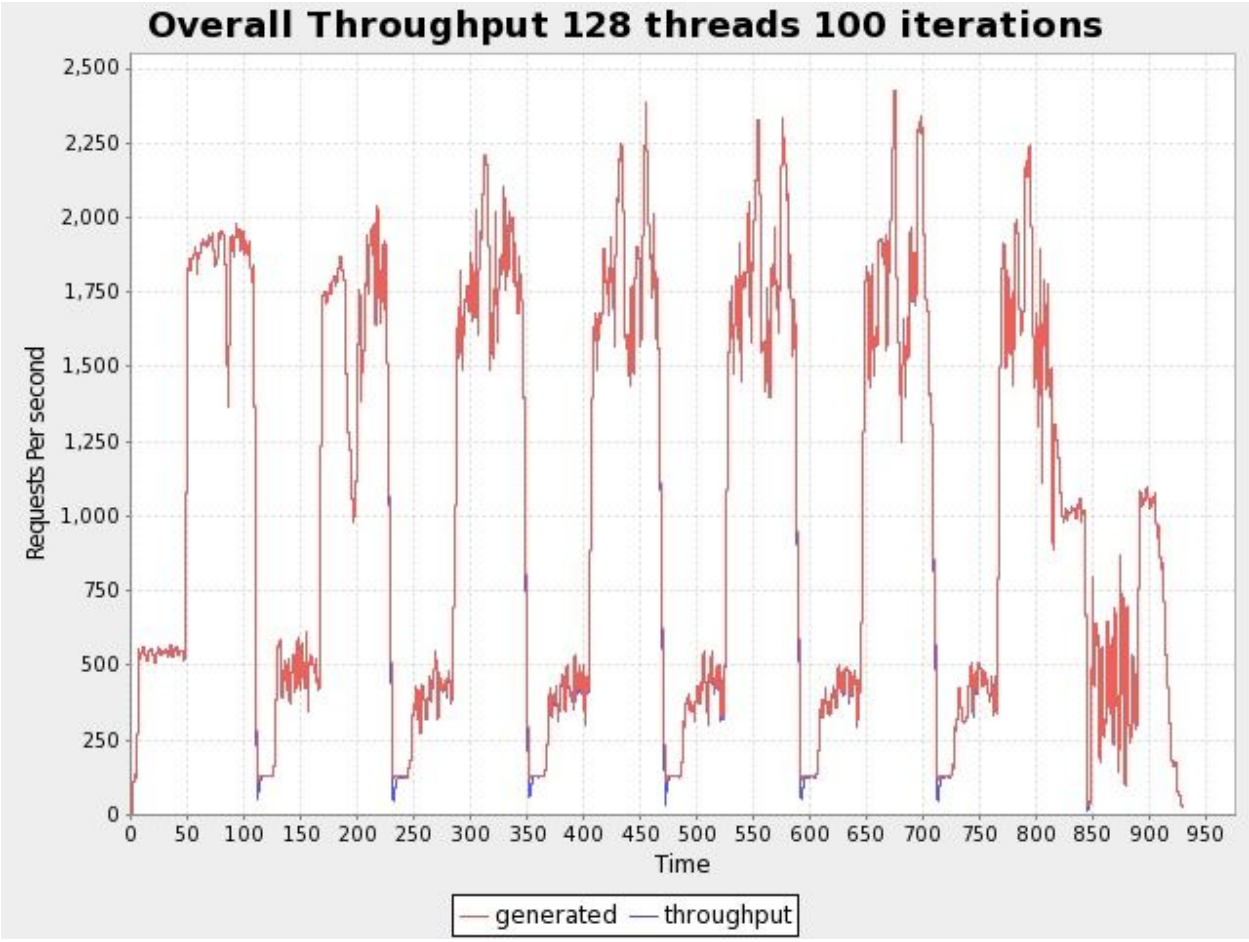
64 Threads



Client starting.... Time: 1542245203435
WARMUP: All threads(6) running....
WARMUP complete: Time 42.294 seconds
LOADING: All threads(32) running....
LOADING complete: Time 82.484 seconds
PEAK: All threads(64) running....

PEAK complete: Time 338.906 seconds
COOLDOWN: All threads(16) running....
COOLDOWN complete: Time 54.911 seconds
==========
Total number of requests sent: 481000
Total number of Successful responses: 480958
Test Wall Time: 518.595 seconds
Overall throughput across all phases: 927.506050000482 rps.
P95 Latency = 38 ms.
P99 Latency = 50 ms.

128 Threads



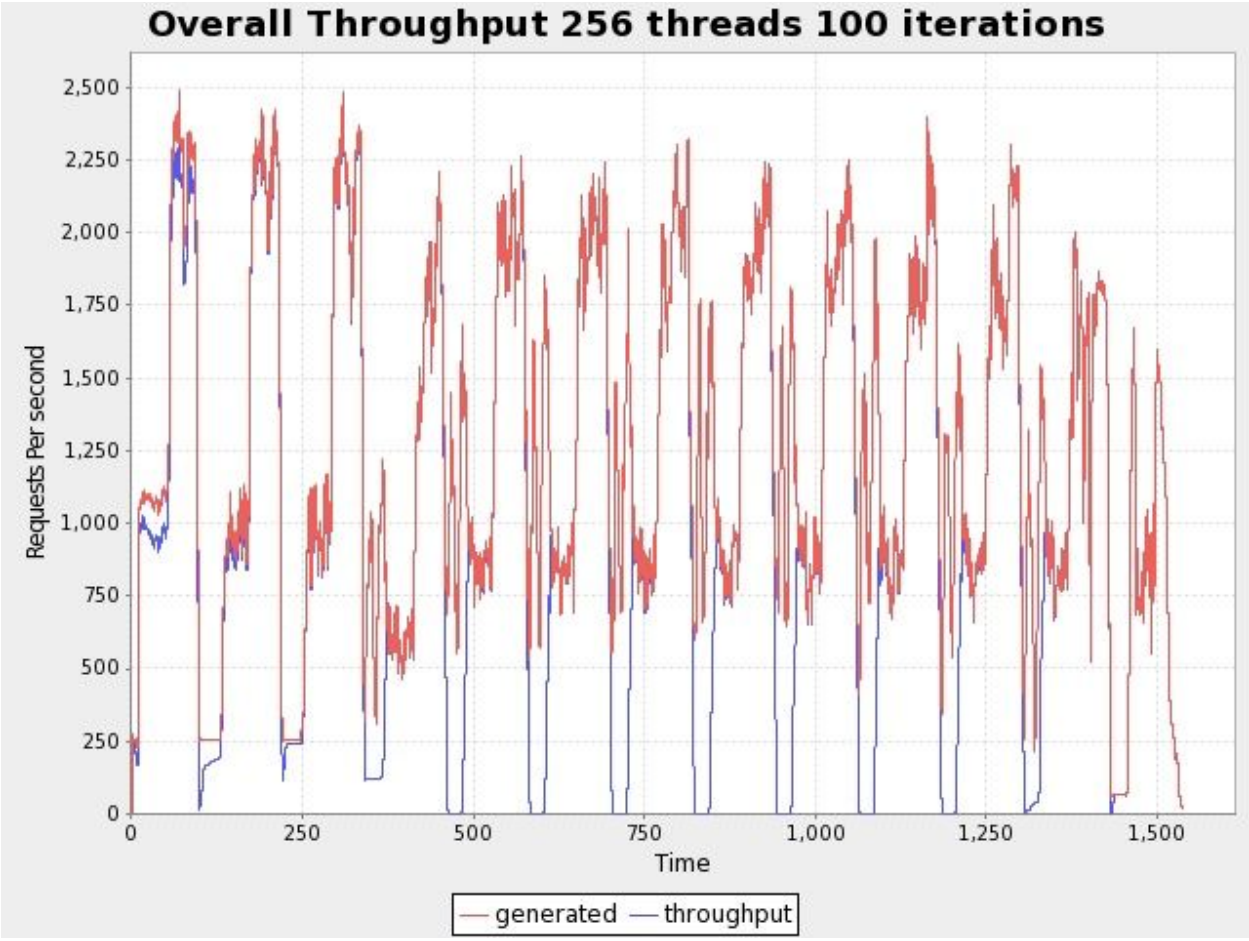
Client starting.... Time: 1542245788281
WARMUP: All threads(12) running....
WARMUP complete: Time 42.826 seconds
LOADING: All threads(64) running....
LOADING complete: Time 142.905 seconds
PEAK: All threads(128) running....
PEAK complete: Time 637.77 seconds

COOLDOWN: All threads(32) running....
COOLDOWN complete: Time 100.332 seconds
==========
Total number of requests sent: 962000
Total number of Successful responses: 959155
Test Wall Time: 923.835 seconds
Overall throughput across all phases: 1041.3114896058278 rps.
P95 Latency = 34 ms.
P99 Latency = 50 ms.

Step 1.2 Compare Results with EC2 from Assignment 2

Lambda Latency is much higher than ec2 load balancer because of cold startup of lambda new tasks rather than minimum tasks already preconfigured to run in load balancer in last assignment. Also, we are not able to reuse db connections as lambda works behind the scene and it is hard to configure higher connection pool size there unless we pre-configure lambda explicitly somehow to use max tasks and reuse tasks for longer time etc.
failure rate is almost similar...
since latency is high, wall clock time of test is also very high...

Step 1.3 >128 THREADS -> 256 Threads

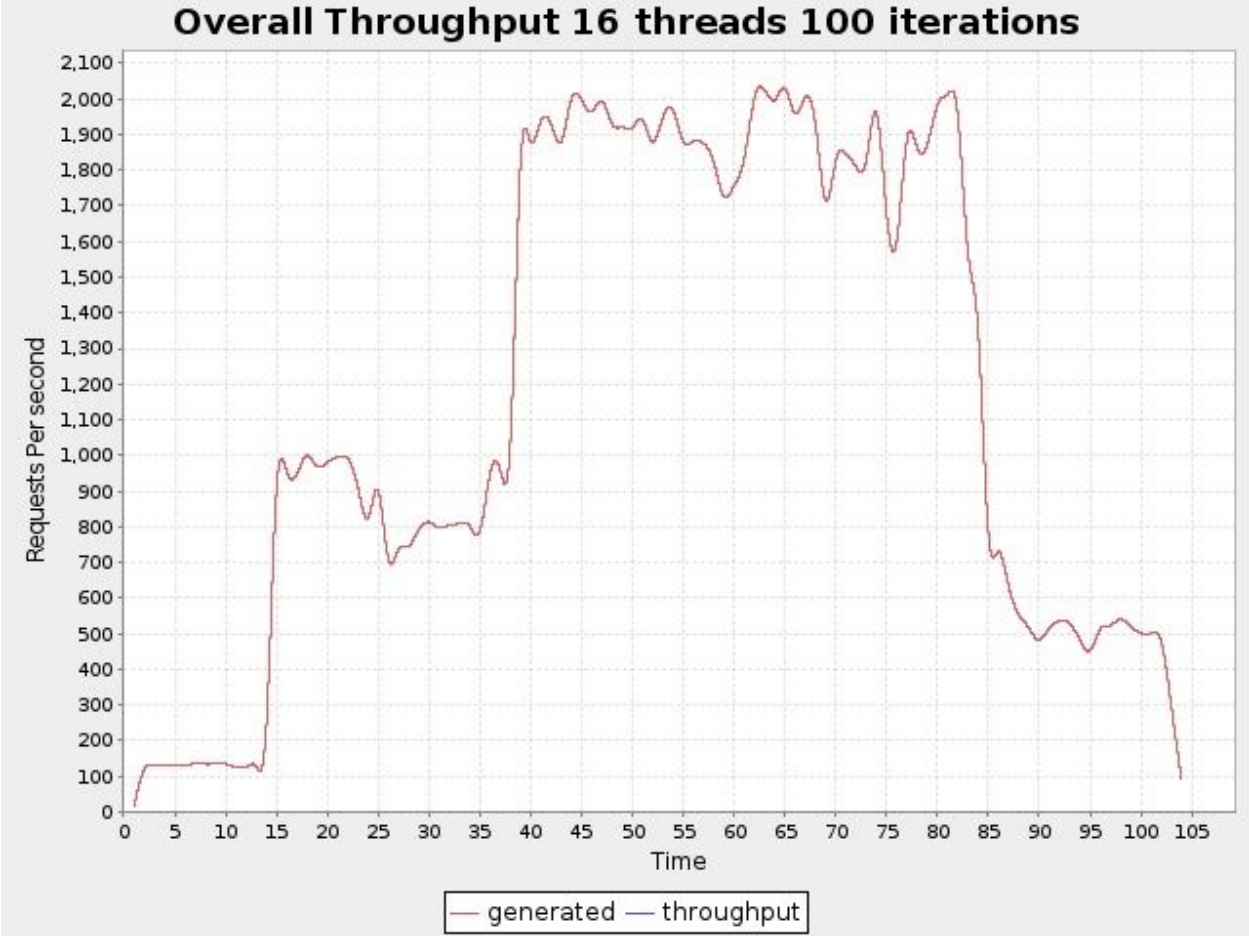


Client starting.... Time: 1542246783319
WARMUP: All threads(25) running....
WARMUP complete: Time 45.235 seconds
LOADING: All threads(128) running....
LOADING complete: Time 283.907 seconds
PEAK: All threads(256) running....
PEAK complete: Time 1069.696 seconds
COOLDOWN: All threads(64) running....
COOLDOWN complete: Time 127.473 seconds
==========
Total number of requests sent: 1925500
Total number of Successful responses: 1627229
Test Wall Time: 1526.311 seconds
Overall throughput across all phases: 1261.5384413792472 rps.
P95 Latency = 66 ms.
P99 Latency = 78 ms.

STEP 2 GCLOUD VM INSTANCE WITH **CLOUD SQL** USING **LOAD BALANCER**

Step 2.1 16, 32, 64, 128, 256 Threads

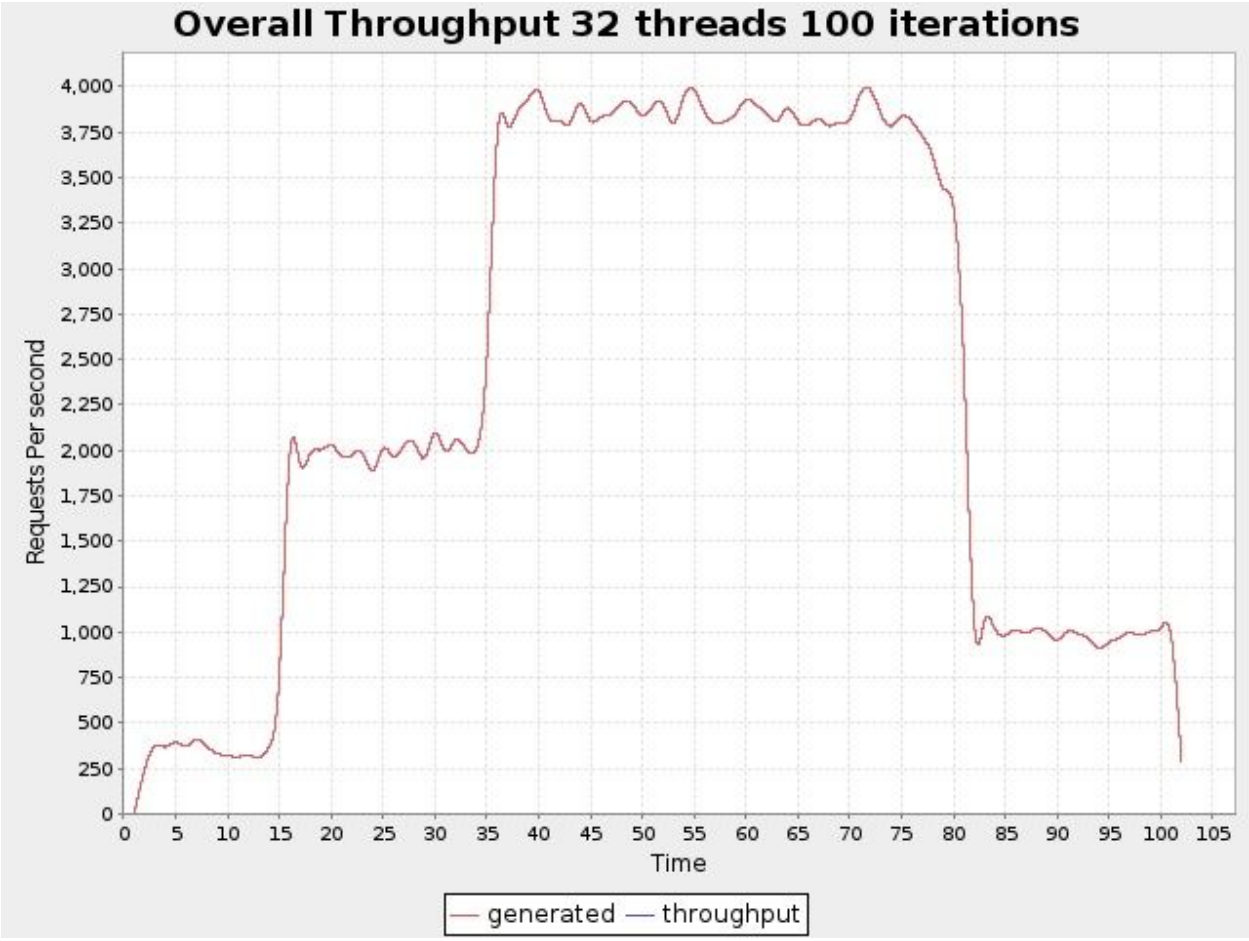
16 Threads



Client starting.... Time: 1542262732270
WARMUP: All threads(1) running....
WARMUP complete: Time 13.123 seconds
LOADING: All threads(8) running....
LOADING complete: Time 24.527 seconds
PEAK: All threads(16) running....
PEAK complete: Time 49.286 seconds
COOLDOWN: All threads(4) running....
COOLDOWN complete: Time 16.058 seconds
==========
Total number of requests sent: 119500
Total number of Successful responses: 119500
Test Wall Time: 102.998 seconds
Overall throughput across all phases: 1160.216703236956 rps.

P95 Latency = 9 ms.
P99 Latency = 11 ms.

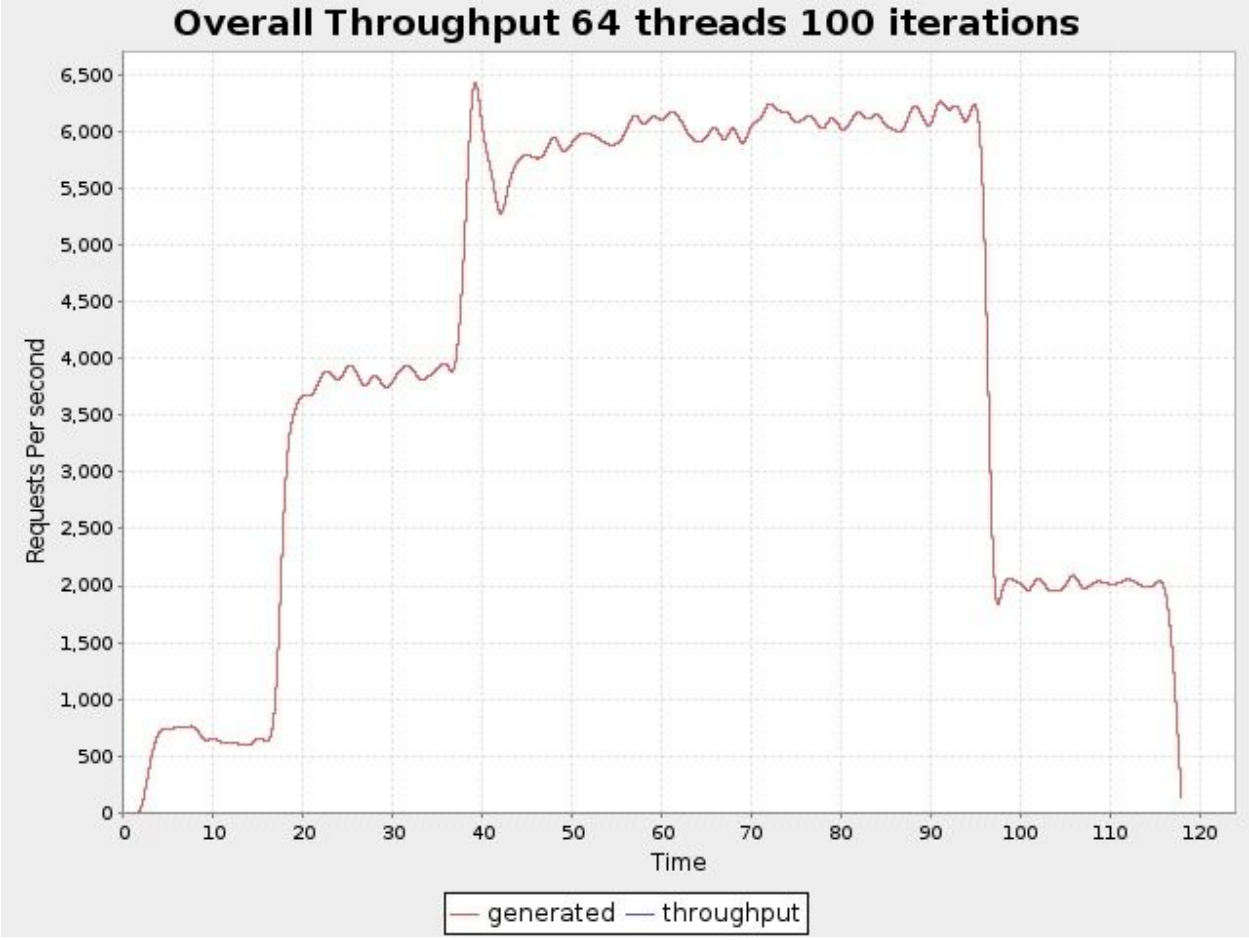
32 Threads



Client starting.... Time: 1542262897831
WARMUP: All threads(3) running....
WARMUP complete: Time 13.768 seconds
LOADING: All threads(16) running....
LOADING complete: Time 20.418 seconds
PEAK: All threads(32) running....
PEAK complete: Time 46.447 seconds
COOLDOWN: All threads(8) running....
COOLDOWN complete: Time 19.751 seconds
==========
Total number of requests sent: 240500
Total number of Successful responses: 240500
Test Wall Time: 100.388 seconds

Overall throughput across all phases: 2395.7046658963222 rps.
P95 Latency = 9 ms.
P99 Latency = 13 ms.

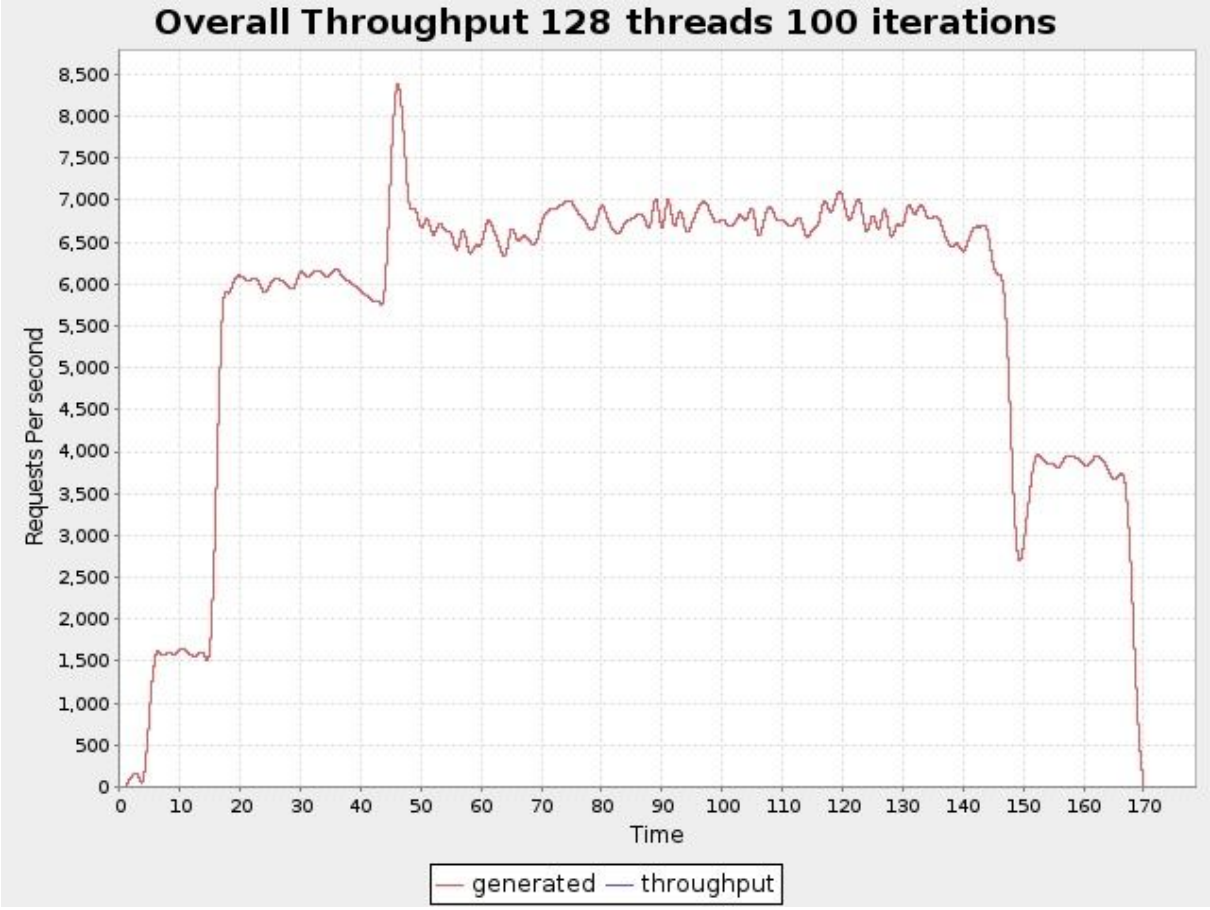
64 Threads



Client starting.... Time: 1542263061850
WARMUP: All threads(6) running....
WARMUP complete: Time 15.061 seconds
LOADING: All threads(32) running....
LOADING complete: Time 21.675 seconds
PEAK: All threads(64) running....
PEAK complete: Time 59.022 seconds
COOLDOWN: All threads(16) running....
COOLDOWN complete: Time 19.855 seconds
==========
Total number of requests sent: 481000
Total number of Successful responses: 481000

Test Wall Time: 115.616 seconds
Overall throughput across all phases: 4160.32383061168 rps.
P95 Latency = 8 ms.
P99 Latency = 10 ms.

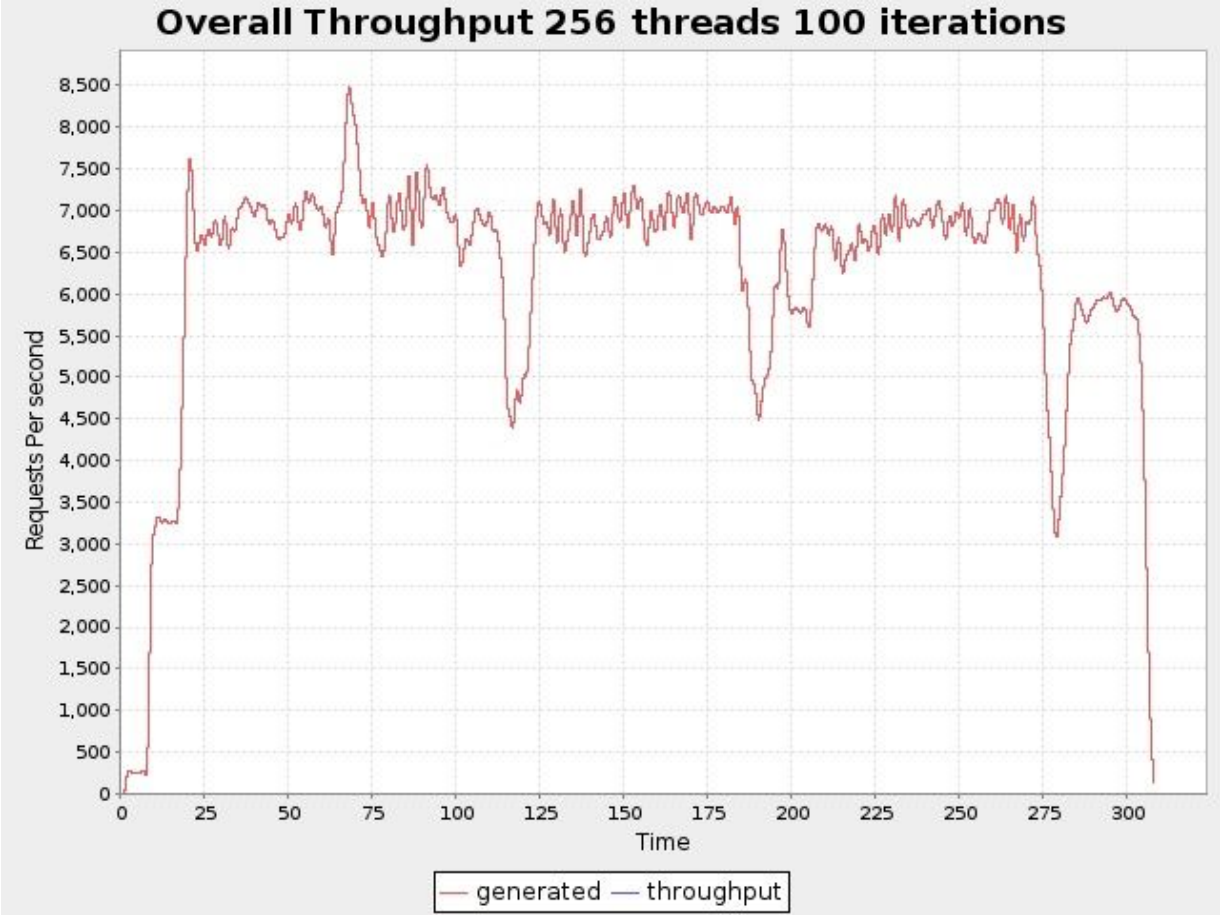
128 Threads



Client starting.... Time: 1542263243228
WARMUP: All threads(12) running....
WARMUP complete: Time 12.469 seconds
LOADING: All threads(64) running....
LOADING complete: Time 30.42 seconds
PEAK: All threads(128) running....
PEAK complete: Time 106.05 seconds
COOLDOWN: All threads(32) running....
COOLDOWN complete: Time 16.591 seconds
==========
Total number of requests sent: 962000
Total number of Successful responses: 962000

Test Wall Time: 165.534 seconds
Overall throughput across all phases: 5811.494919472737 rps.
P95 Latency = 10 ms.
P99 Latency = 13 ms.

256 Threads



Client starting.... Time: 1542263478509
WARMUP: All threads(25) running....
WARMUP complete: Time 12.627 seconds
LOADING: All threads(128) running....
LOADING complete: Time 50.262 seconds
PEAK: All threads(256) running....
PEAK complete: Time 213.147 seconds
COOLDOWN: All threads(64) running....
COOLDOWN complete: Time 23.741 seconds
==========
Total number of requests sent: 1925500

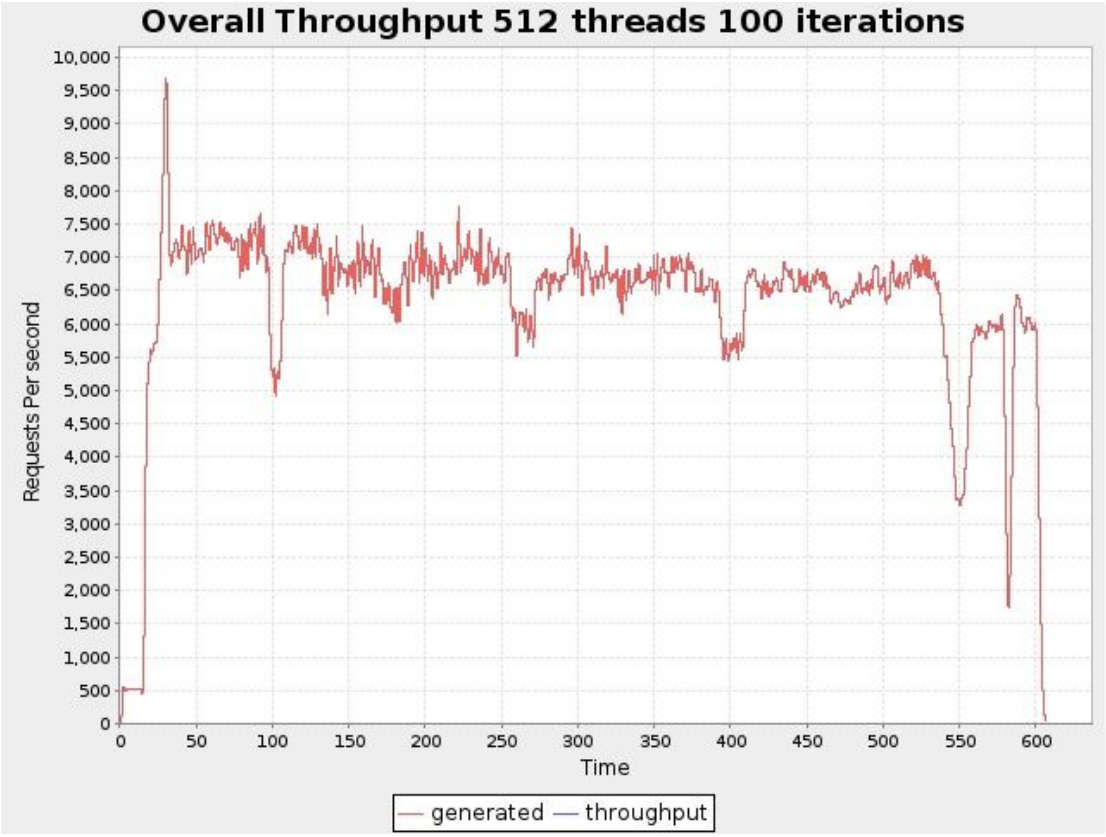
Total number of Successful responses: 1925500
Test Wall Time: 299.779 seconds
Overall throughput across all phases: 6423.064991210192 rps.
P95 Latency = 10 ms.
P99 Latency = 18 ms.

Step 2.2 Compare results with ec2 from Assignment2

gc2 load balancer (east region) test when run in east region gc compute machine considerably super high requests per second (6-7x) and super low latency (10x low) as network path is very cheap from within one data center (or maybe within same subnet) and doesn't have to travel over so many hops to/from client.

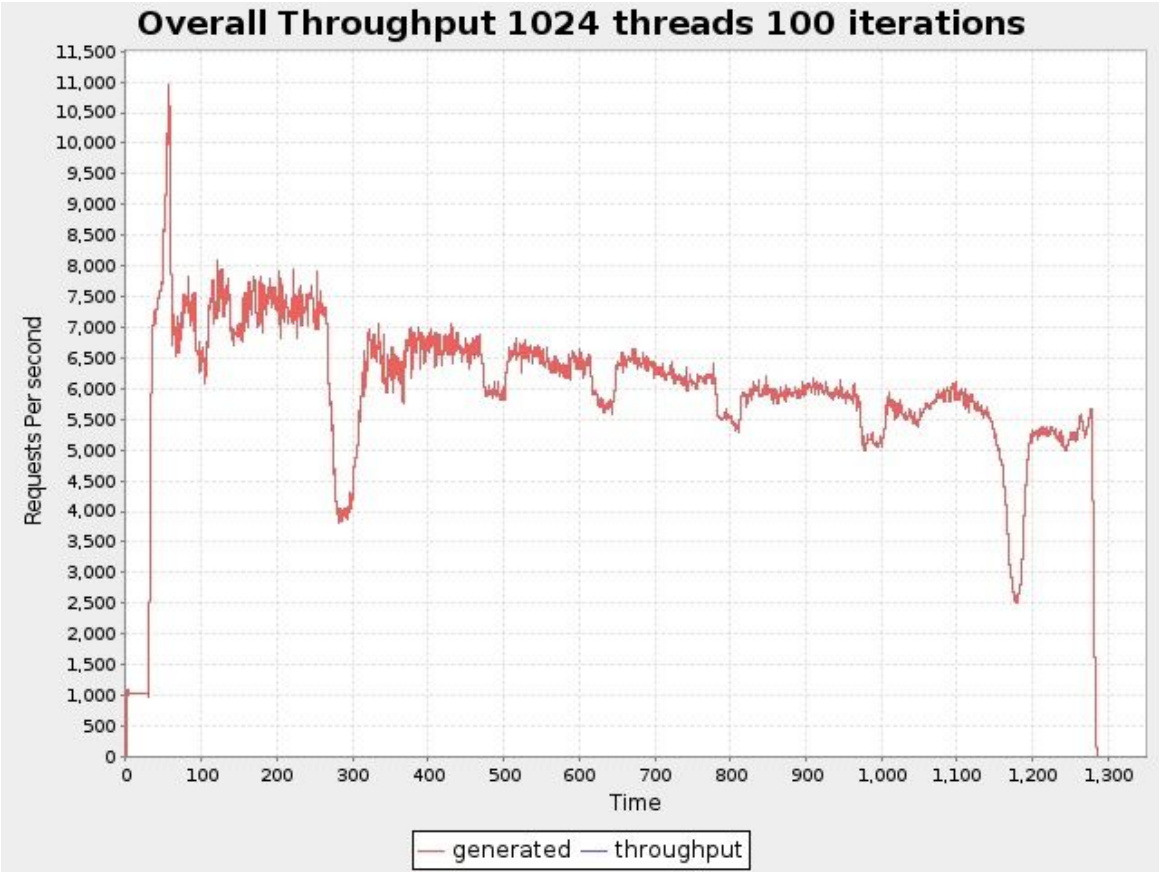
gc2 load balancer (east region) test when run from laptop (in west region)
since laptop is running in west it has to go from so many extra hops to reach gc2 machines running in east cose so latency is 1.5-1.7x higher than ec2 load balancer test.similar request per seocnd is 20% lower than ec2 load balancer... but otherwise graph is almost similar just stretched over time axis.

Step 2.3 >256 Threads (512 and 1024 threads)
512 Threads



Client starting.... Time: 1542263856647
WARMUP: All threads(51) running....
WARMUP complete: Time 15.327 seconds
LOADING: All threads(256) running....
LOADING complete: Time 105.893 seconds
PEAK: All threads(512) running....
PEAK complete: Time 426.247 seconds
COOLDOWN: All threads(128) running....
COOLDOWN complete: Time 42.382 seconds
==========
Total number of requests sent: 3852500
Total number of Successful responses: 3852496
Test Wall Time: 589.883 seconds
Overall throughput across all phases: 6530.95613875972 rps.
P95 Latency = 8 ms.
P99 Latency = 10 ms.

1024 Threads



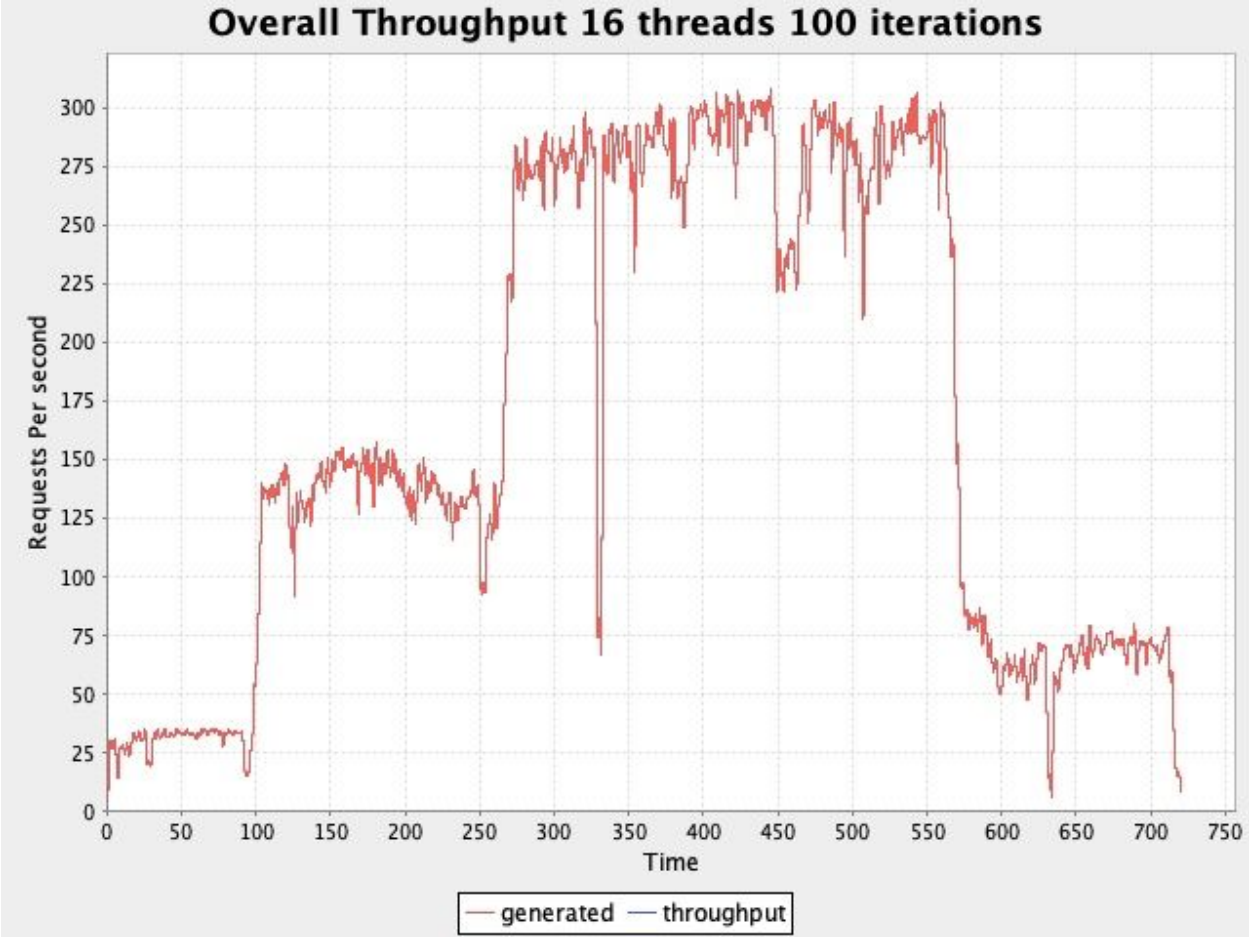
Client starting.... Time: 1542264541473
WARMUP: All threads(102) running....
WARMUP complete: Time 27.929 seconds
LOADING: All threads(512) running....
LOADING complete: Time 224.572 seconds
PEAK: All threads(1024) running....
PEAK complete: Time 918.471 seconds
COOLDOWN: All threads(256) running....
COOLDOWN complete: Time 85.043 seconds
==========
Total number of requests sent: 7705000
Total number of Successful responses: 7704994
Test Wall Time: 1256.02 seconds
Overall throughput across all phases: 6134.456457699718 rps.
P95 Latency = 12 ms.
P99 Latency = 14 ms.

NOTE : I got the above results when I ran on my Lambda and Google Cloud Instance. Because I was having some network issues, But I

also completed the tasks through my laptop. (Google cloud instance is in zone us-east as west did not have enough instances and that is why time taken can be more).

LAMBDA

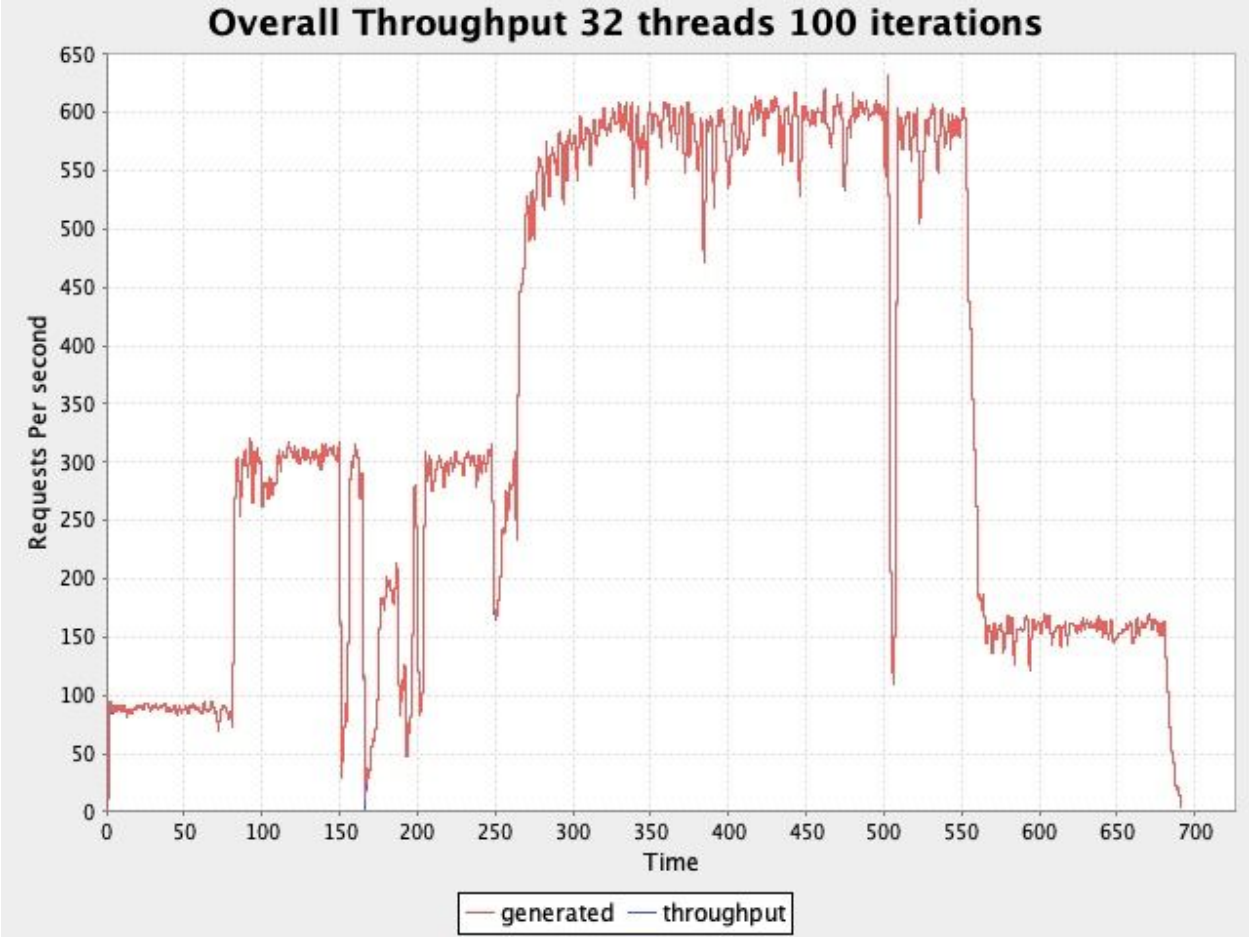
Lambda 16 Threads:



Client starting.... Time: 1542401178857
WARMUP: All threads(1) running....
WARMUP complete: Time 90.795 seconds
LOADING: All threads(8) running....
LOADING complete: Time 158.109 seconds
PEAK: All threads(16) running....
PEAK complete: Time 324.834 seconds
COOLDOWN: All threads(4) running....
COOLDOWN complete: Time 145.247 seconds
==========
Total number of requests sent: 119500
Total number of Successful responses: 119500

Test Wall Time: 718.986 seconds
Overall throughput across all phases: 166.20629608921453 rps.
P95 Latency = 85 ms.
P99 Latency = 95 ms.

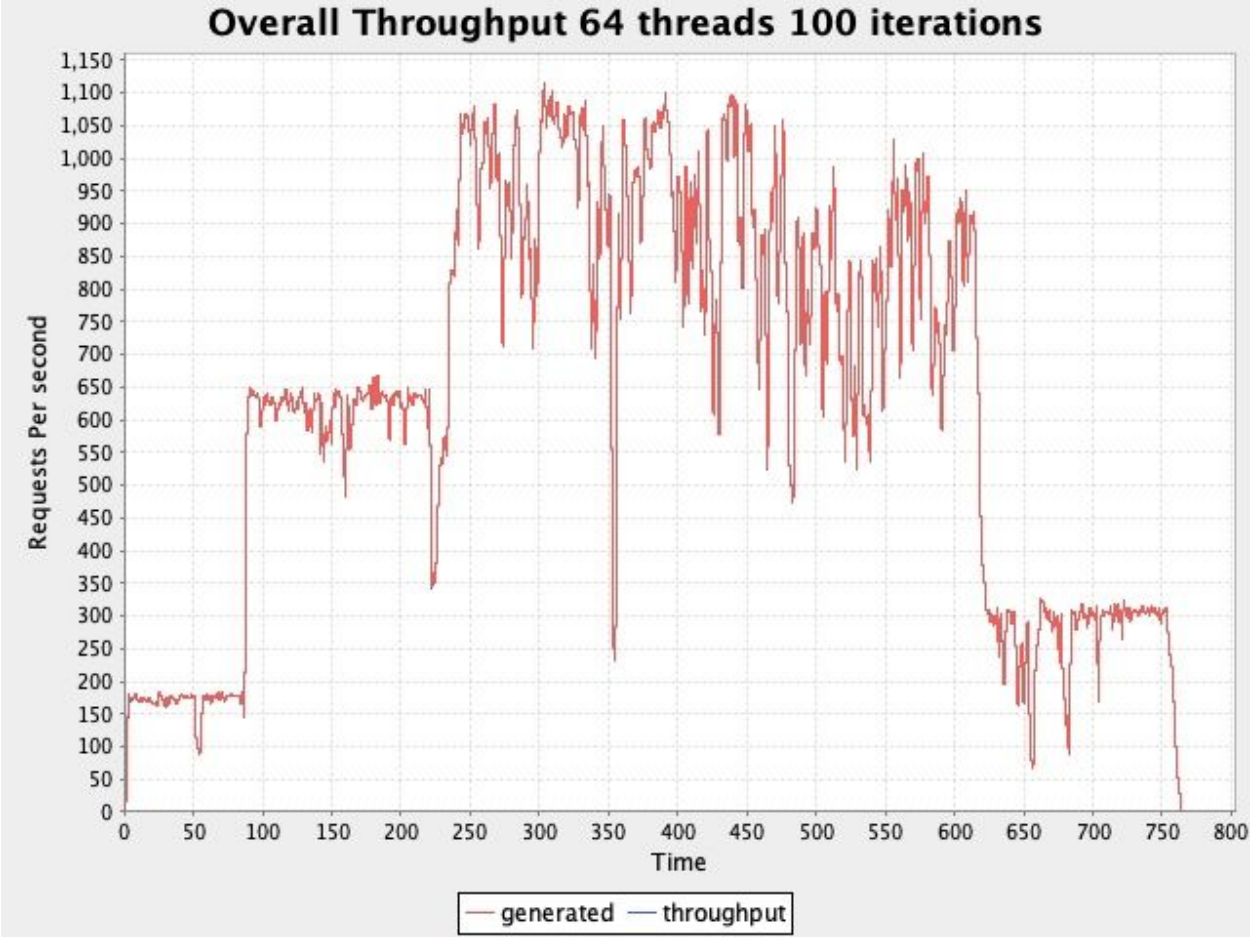
Client starting.... Time: 1542401962418
WARMUP: All threads(3) running....
Lambda 32 Threads :



WARMUP complete: Time 76.148 seconds
LOADING: All threads(16) running....
LOADING complete: Time 1125.806 seconds
PEAK: All threads(32) running....
PEAK complete: Time 316.572 seconds
COOLDOWN: All threads(8) running....
COOLDOWN complete: Time 125.27 seconds
==========
Total number of requests sent: 240500

Total number of Successful responses: 240479
Test Wall Time: 1643.798 seconds
Overall throughput across all phases: 146.30751467029404 rps.
P95 Latency = 68 ms.
P99 Latency = 74 ms.

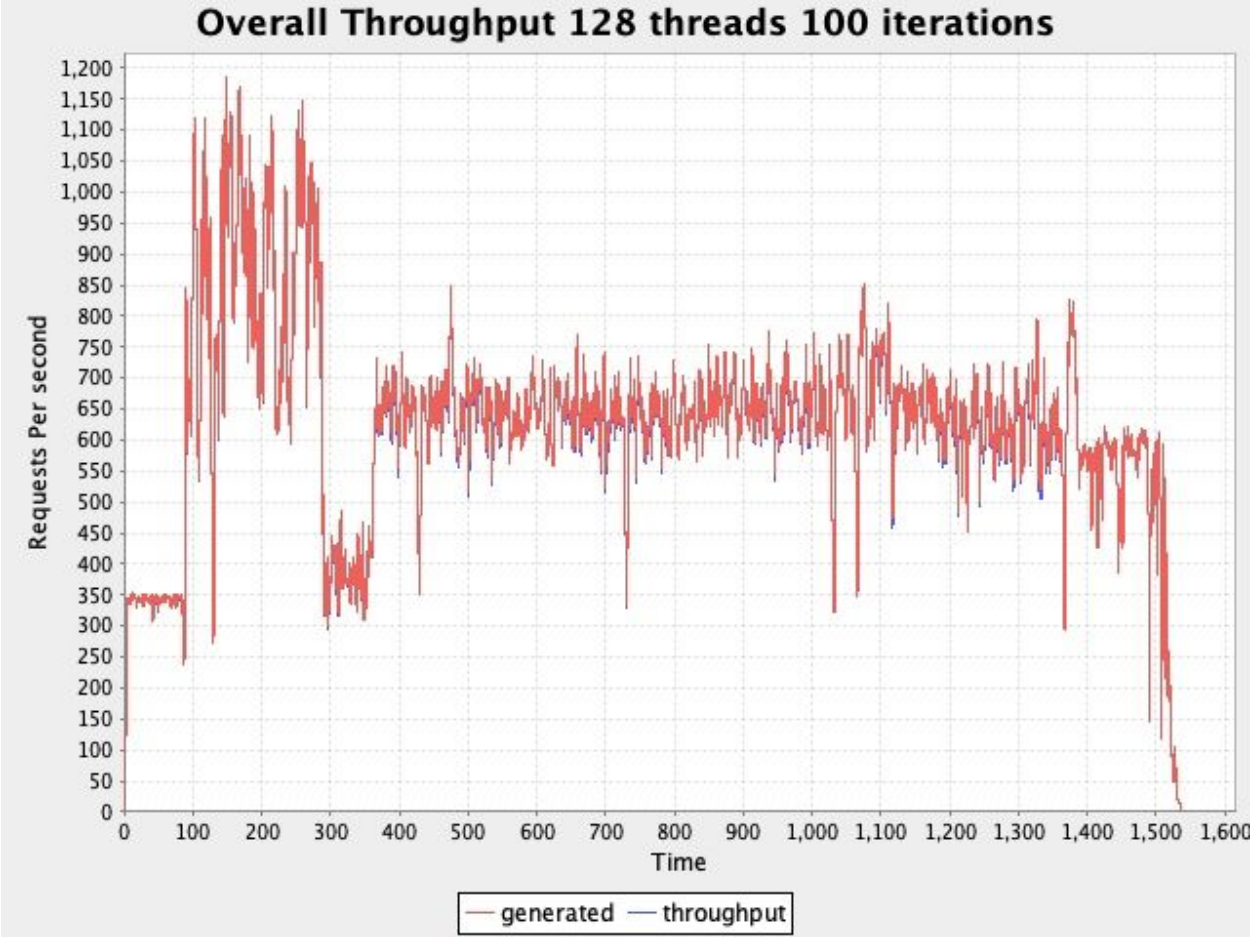
Lambda 64 Threads :



Client starting.... Time: 1542403671432
WARMUP: All threads(6) running....
WARMUP complete: Time 81.44 seconds
LOADING: All threads(32) running....
LOADING complete: Time 138.256 seconds
PEAK: All threads(64) running....
PEAK complete: Time 405.074 seconds
COOLDOWN: All threads(16) running....
COOLDOWN complete: Time 136.985 seconds
==========

Total number of requests sent: 481000
Total number of Successful responses: 480946
Test Wall Time: 761.755 seconds
Overall throughput across all phases: 631.4366167599819 rps.
P95 Latency = 62 ms.
P99 Latency = 65 ms.

Lambda 128 Threads :



Client starting.... Time: 1542404500223
WARMUP: All threads(12) running....
WARMUP complete: Time 80.42 seconds
LOADING: All threads(64) running....
LOADING complete: Time 204.405 seconds
PEAK: All threads(128) running....
PEAK complete: Time 1099.829 seconds
COOLDOWN: All threads(32) running....
COOLDOWN complete: Time 148.669 seconds
==========

Total number of requests sent: 962000
Total number of Successful responses: 958690
Test Wall Time: 1533.326 seconds
Overall throughput across all phases: 627.3943049292844 rps.
P95 Latency = 296 ms.
P99 Latency = 317 ms.

GOOGLE CLOUD

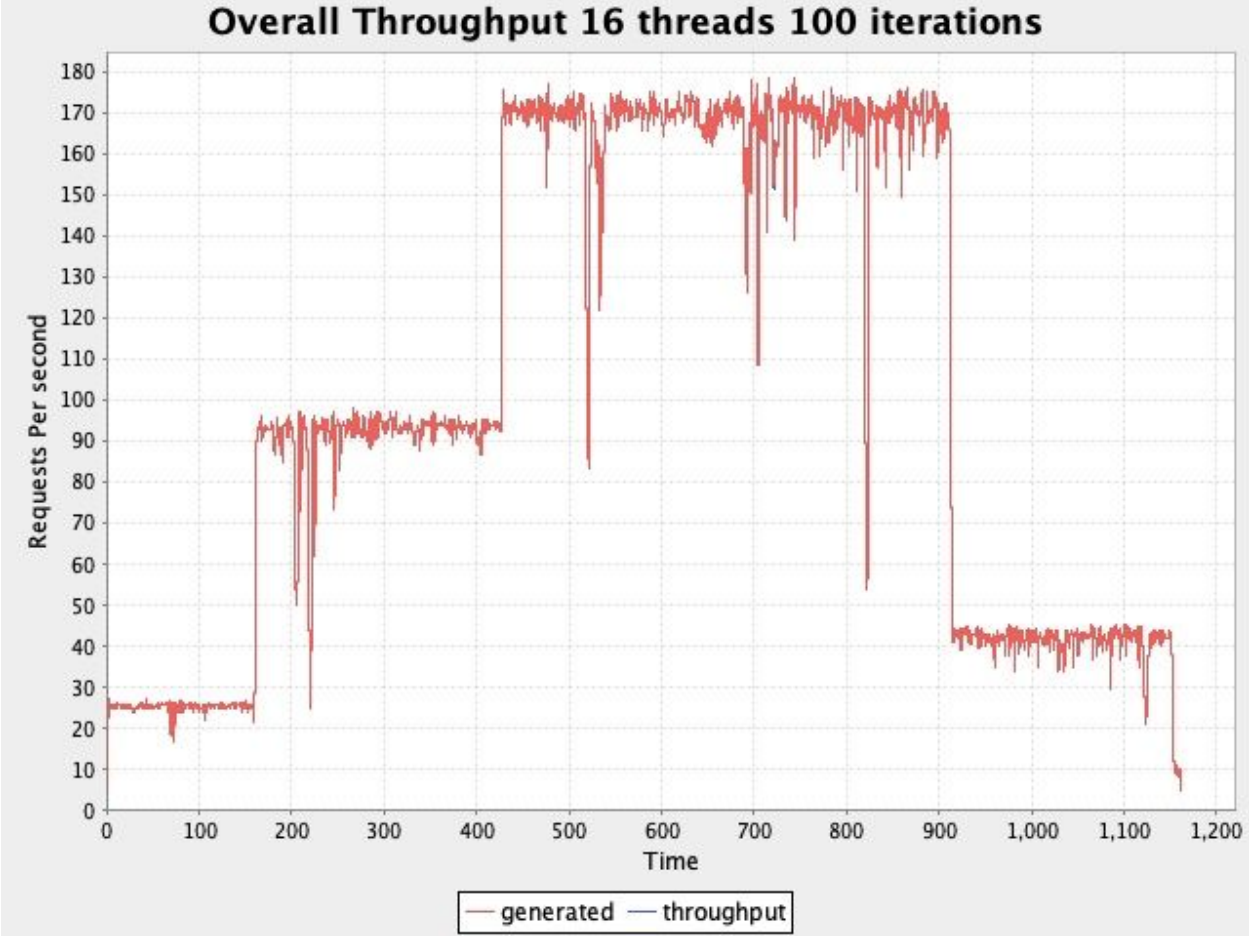
Google Cloud 16 Threads

Google Cloud CMD :
I used Bash Script to run Google Cloud Load Balancer for 16, 32, 64, 128, 256 threads.

Script I used :

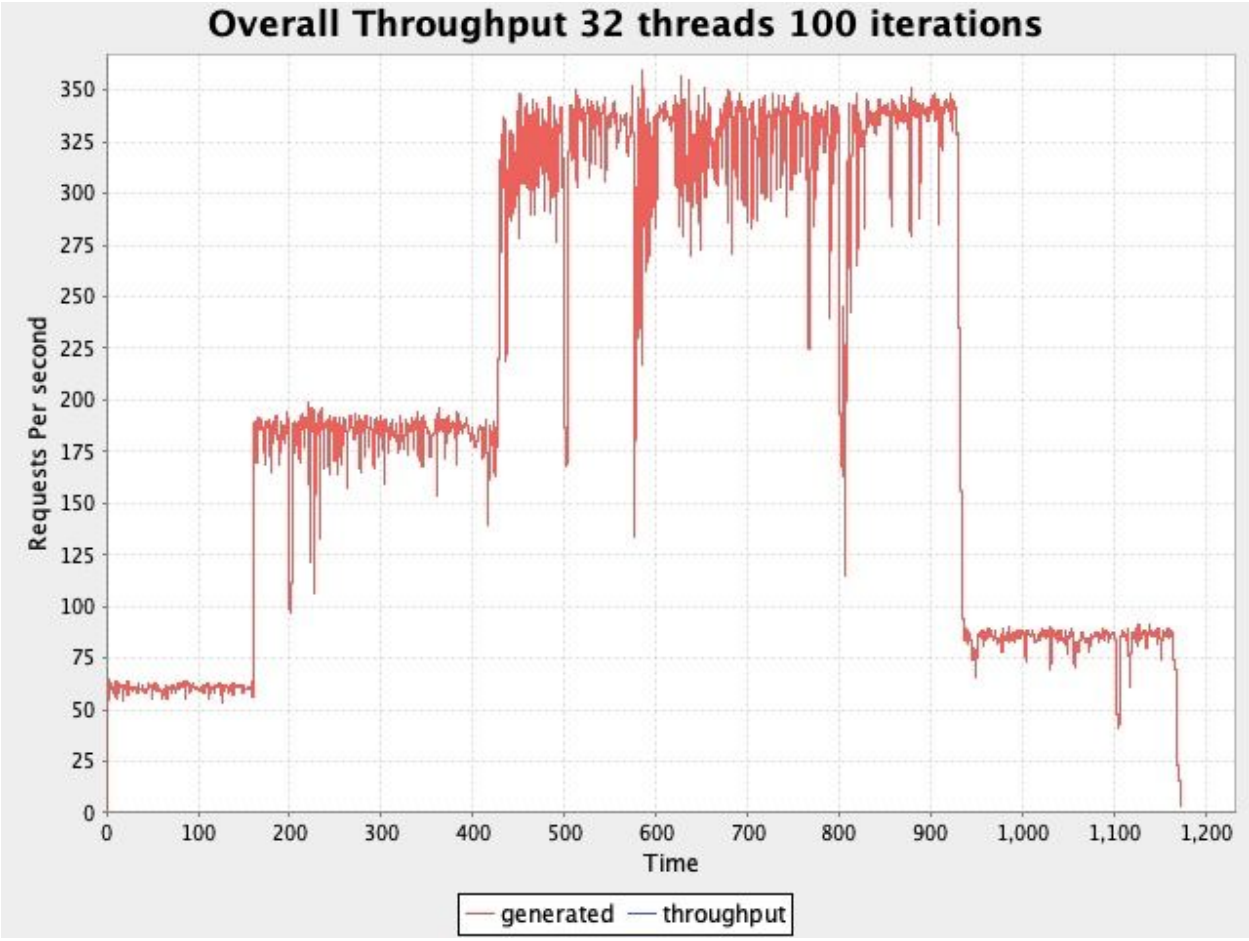
```
#!/bin/bash

for num_threads in 16 32 64 128 256 ; do
        echo "running for $num_threads threads"
        java -jar bsdsassignment2-stress-jar-with-dependencies.jar -h "35.244.213.164" -p
8080 -u "/bsdsassignment2-webapp/stepcounter" -t "$num_threads" -g
"photos/StepCounterThroughput_googleCloud_${num_threads}.jpeg" 2>&1 | tee
"results/stats_googleCloud_${num_threads}.txt"
        echo "finished for $num_threads threads"
        sleep 60
Done
```

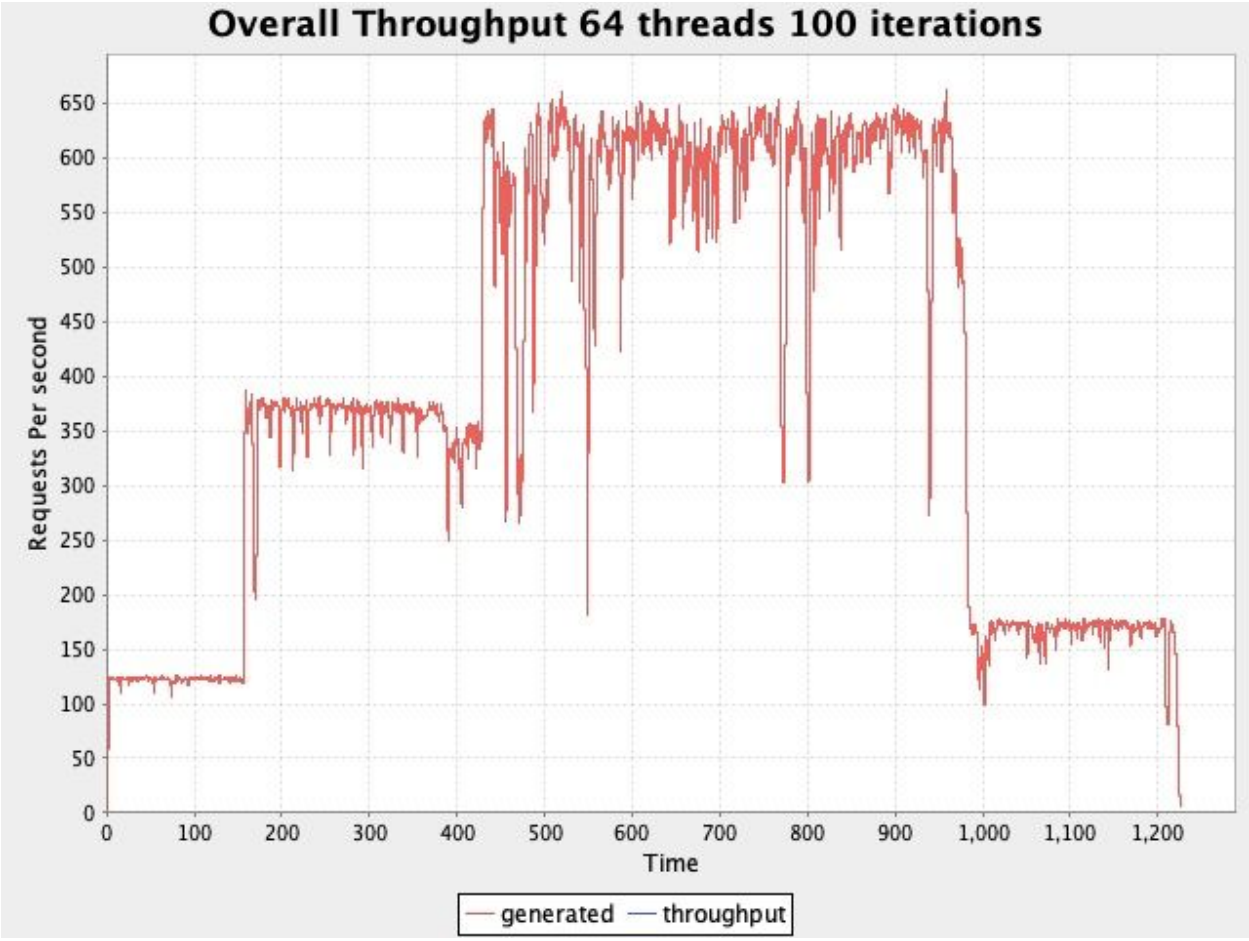


Client starting.... Time: 1542345451635
WARMUP: All threads(1) running....
WARMUP complete: Time 144.717 seconds
LOADING: All threads(8) running....
LOADING complete: Time 241.575 seconds
PEAK: All threads(16) running....
PEAK complete: Time 533.179 seconds
COOLDOWN: All threads(4) running....
COOLDOWN complete: Time 240.5 seconds
==========
Total number of requests sent: 119500
Total number of Successful responses: 119498
Test Wall Time: 1159.974 seconds
Overall throughput across all phases: 103.01955043819949 rps.
P95 Latency = 123 ms.
P99 Latency = 131 ms.

Google Cloud 32 Threads

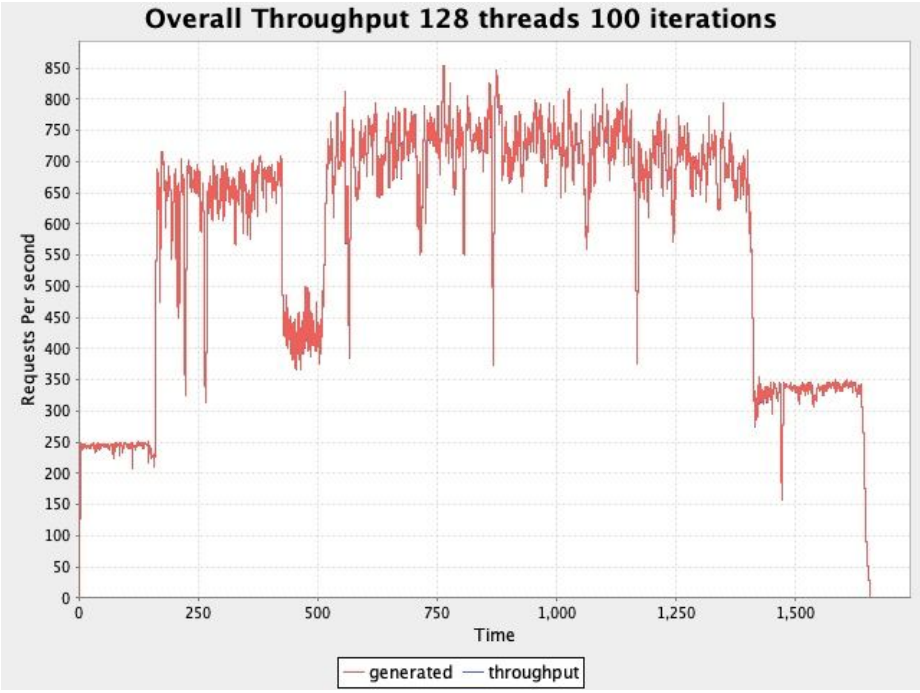


Client starting.... Time: 1542346675623
WARMUP: All threads(3) running....
WARMUP complete: Time 144.32 seconds
LOADING: All threads(16) running....
LOADING complete: Time 241.972 seconds
PEAK: All threads(32) running....
PEAK complete: Time 548.14 seconds
COOLDOWN: All threads(8) running....
COOLDOWN complete: Time 237.082 seconds
==========
Total number of requests sent: 240500
Total number of Successful responses: 240500
Test Wall Time: 1171.517 seconds
Overall throughput across all phases: 205.28938120402862 rps.
P95 Latency = 106 ms.
P99 Latency = 109 ms.
Google Cloud 64 Threads



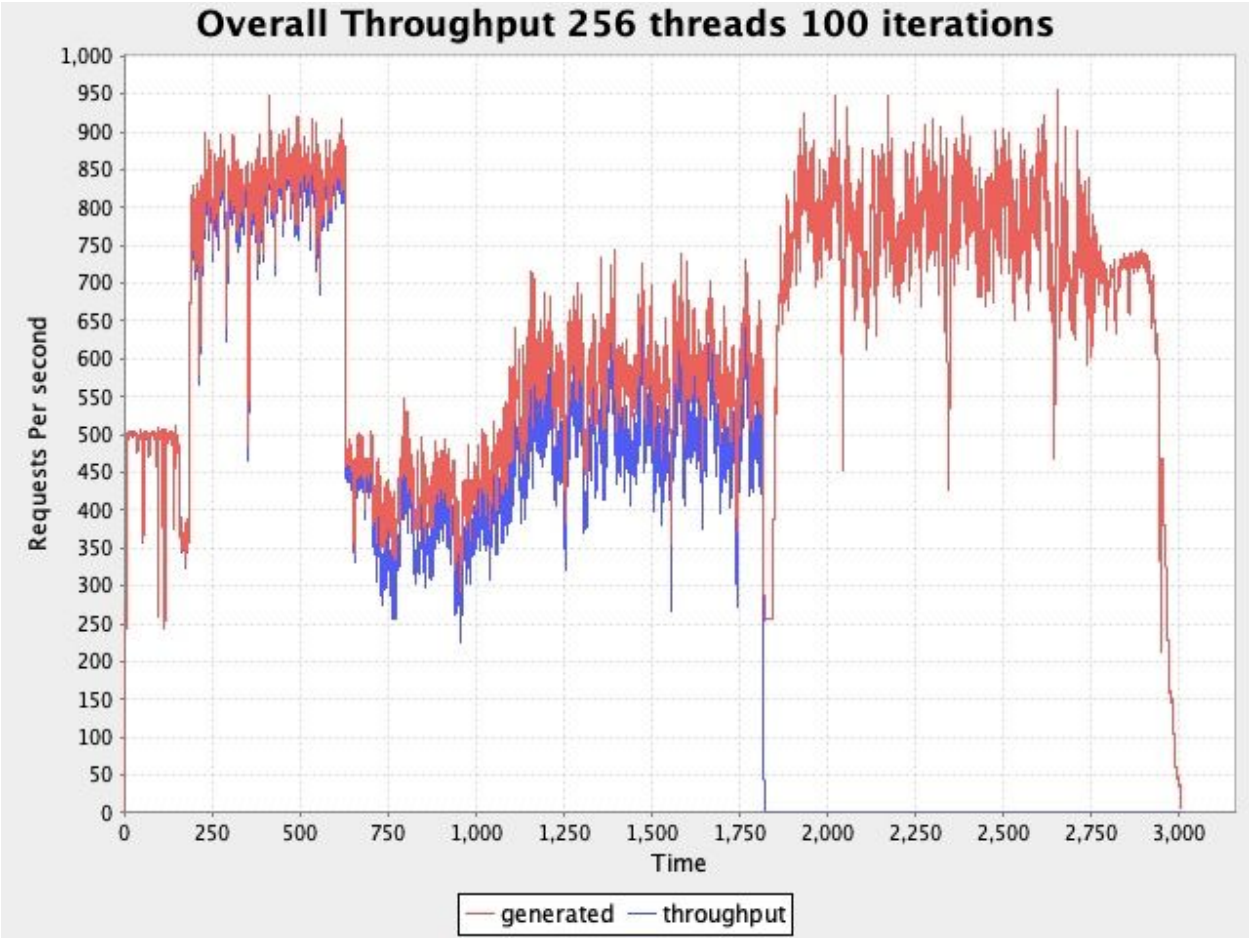
Client starting.... Time: 1542347912749
WARMUP: All threads(6) running....
WARMUP complete: Time 140.633 seconds
LOADING: All threads(32) running....
LOADING complete: Time 241.949 seconds
PEAK: All threads(64) running....
PEAK complete: Time 601.884 seconds
COOLDOWN: All threads(16) running....
COOLDOWN complete: Time 239.053 seconds
==========
Total number of requests sent: 481000
Total number of Successful responses: 480998
Test Wall Time: 1223.522 seconds
Overall throughput across all phases: 393.12738144471456 rps.
P95 Latency = 217 ms.
P99 Latency = 231 ms.

Google Cloud 128 Threads



Client starting.... Time: 1542349203468
WARMUP: All threads(12) running....
WARMUP complete: Time 142.126 seconds
LOADING: All threads(64) running....
LOADING complete: Time 280.967 seconds
PEAK: All threads(128) running....
PEAK complete: Time 999.765 seconds
COOLDOWN: All threads(32) running....
COOLDOWN complete: Time 229.693 seconds
==========
Total number of requests sent: 962000
Total number of Successful responses: 961730
Test Wall Time: 1652.554 seconds
Overall throughput across all phases: 582.1292375317236 rps.
P95 Latency = 108 ms.
P99 Latency = 111 ms.

Google Cloud 256 Threads :



Client starting.... Time: 1542350925626
WARMUP: All threads(25) running....
WARMUP complete: Time 149.381 seconds
LOADING: All threads(128) running....
LOADING complete: Time 475.555 seconds
PEAK: All threads(256) running....
PEAK complete: Time 2160.768 seconds
COOLDOWN: All threads(64) running....
COOLDOWN complete: Time 219.21 seconds
==========
Total number of requests sent: 1925500
Total number of Successful responses: 983757
Test Wall Time: 3004.918 seconds
Overall throughput across all phases: 640.7828766042867 rps.
P95 Latency = 0 ms.
P99 Latency = 0 ms